Additional assumption needed for confidence intervals & hypothesis tests: Stat 564 10-3-17

(1)

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent, (i.i.d.) identically distributed $N(0,\sigma^2)$ random variables.

Facts: (1) $\hat{\beta}_1$ is normally distributed

(2) $\frac{(n-2)\hat{\sigma}^2}{\sigma^2}$ has a χ^2 distribution with $n-2$ df

(3) $\hat{\beta}_1$ and $\hat{\sigma}^2$ are independent

(2)

Use these facts to get:

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \sim N(0,1)$$

Recall: If $Z \sim N(0,1)$ and $W \sim \chi^2_\nu$ and Z, W are independent, then

$$Z \Big/ \sqrt{\frac{W}{\nu}} \sim t_\nu$$

3

$$\frac{\dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\cancel{\sigma^2}}{S_{xx}}}}}{\sqrt{\dfrac{(n-2)\hat{\sigma}^2}{\cancel{\sigma^2}} \Big/ \cancel{(n-2)}}} \sim t_{n-2}$$

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}} \sim t_{n-2}$$

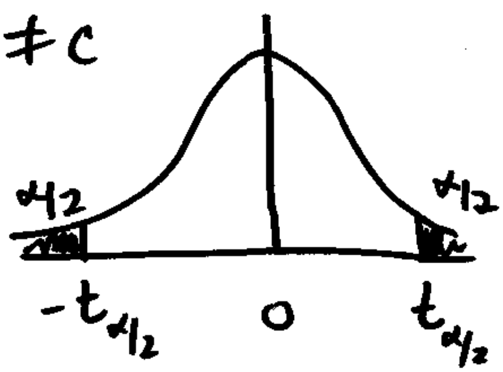
4

<u>Hypothesis test</u> $\quad H_0: \beta_1 = c$

$\quad H_1: \beta_1 \neq c$

Test statistic =

$$\frac{\hat{\beta}_1 - c}{\sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}}}$$



and reject H_0 if the test stat. exceeds $t_{\alpha/2}$ in absolute value

Confidence interval for β_1 is ⑤

$$\hat{\beta}_1 \pm t_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

Observe $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
and fit the least squares line

At a new x value, called x_0, observe
k new values of y.
Let $\bar{y}_0$ denote the average of these new y-values.

Consider the random variable $\hat{y}_0 - \bar{y}_0$ ⑥

$$E(\hat{y}_0 - \bar{y}_0) = E\left(\hat{\beta}_0 + \hat{\beta}_1 x_0 - (\beta_0 + \beta_1 x_0 + \bar{\varepsilon}_k)\right)$$

$$= \beta_0 + \beta_1 x_0 - \beta_0 - \beta_1 x_0 - 0$$

$$= 0$$

$$V(\hat{y}_0 - \bar{y}_0) = V(\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0 - \bar{\varepsilon}_k)$$

$$= V(\hat{\beta}_0 + \hat{\beta}_1 x_0 - \bar{\varepsilon}_k)$$

$$= V(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_0 - \bar{\varepsilon}_k)$$

(7)

$$= V\big(\underbrace{\bar{y} - \hat{\beta}_1(\bar{x} - x_0)} - \underbrace{\bar{\varepsilon}_K}\big)$$

$$= V\big(\bar{y} - \hat{\beta}_1(\bar{x} - x_0)\big) + V(\bar{\varepsilon}_K)$$

$$+ \cancel{2Cov}$$

$$= V\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{S_{xy}}{S_{xx}}(\bar{x} - x_0)\right) + \frac{\sigma^2}{K}$$

$$= V\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \left[\sum_{i=1}^{n}(x_i - \bar{x})y_i\right]\frac{\bar{x} - x_0}{S_{xx}}\right) + \frac{\sigma^2}{K}$$

$$= V\left[\sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x} - x_0}{S_{xx}}(x_i - \bar{x})\right)y_i\right] + \frac{\sigma^2}{K}$$

(8)

$$= \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x} - x_0}{S_{xx}}(x_i - \bar{x})\right)^2\sigma^2 + \frac{\sigma^2}{K}$$

$$= \sigma^2\Bigg[\frac{n}{n^2} + \frac{(\bar{x} - x_0)^2}{S_{xx}^2}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$- \frac{2}{n}\frac{\bar{x} - x_0}{S_{xx}}\cancelto{0}{\sum_{i=1}^{n}(x_i - \bar{x})} + \frac{1}{K}\Bigg]$$

$$= \sigma^2\left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}} + \frac{1}{K}\right]$$

$$= V(\hat{y}_0 - \bar{y}_0)$$

Consequences: (9)

(1) Prediction interval for a single new y-value at $x = x_0$ $(k=1)$

$$\hat{y}_0 \pm t_{\alpha/2}\,\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}$$

(2) Confidence interval for the mean of all possible y-values at $x = x_0$ $(k=\infty)$

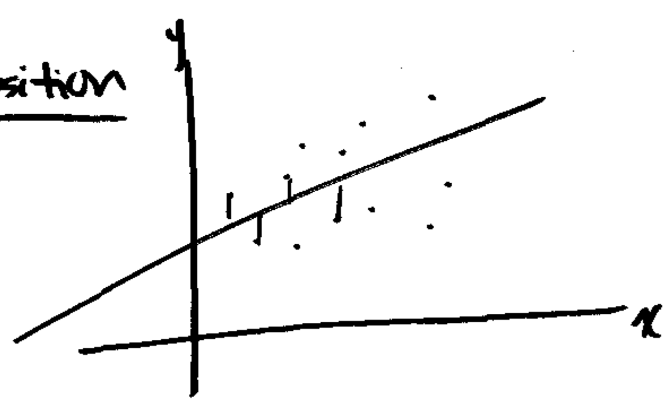
$$\hat{y}_0 \pm t_{\alpha/2}\,\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}$$

(3) Confidence interval for β_0 (10)

$(x_0=0, k=\infty)$

$$\hat{\beta}_0 \pm t_{\alpha/2}\,\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

Variance decomposition



11

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$= \sum_{i=1}^{n}(\underbrace{y_i - \hat{y}_i}+\underbrace{\hat{y}_i - \bar{y}})^2$$

$$= \sum_{i=1}^{n}\overset{(1)}{(y_i - \hat{y}_i)^2} + \sum_{i=1}^{n}\overset{(2)}{(\hat{y}_i - \bar{y})^2}$$

$$+ 2\sum_{i=1}^{n}\overset{(3)}{(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})}$$

(1): SSE, defined last time

12

(2) $\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2$

$$= \sum_{i=1}^{n}(\cancel{\bar{y}} - \hat{\beta}_1\bar{x} + \hat{\beta}_1 x_i - \cancel{\bar{y}})^2$$

$$= \sum_{i=1}^{n}\left[\hat{\beta}_1(x_i - \bar{x})\right]^2 = \hat{\beta}_1^2 \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})^2}_{S_{xx}}$$

$$= \left[\frac{S_{xy}}{S_{xx}}\right]^2 S_{xx} = \frac{S_{xy}^2}{S_{xx}}$$

13

Recall from last time: $SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$

$\Rightarrow S_{yy} = SSE + \frac{S_{xy}^2}{S_{xx}}$

$\Rightarrow$ ③ $= 0$

$$\therefore \quad S_{yy} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

$$\underset{\text{total}}{SST} = \underset{\text{error}}{SSE} + \underset{\text{regression}}{SSR}$$

14

Definition:

$R^2 = \frac{SSR}{SST}$ = proportion of the variation in y that is explained by x.

$0 \le R^2 \le 1$

0: x, as a linear predictor, has no effect on y

1: perfect fit of line

$R^2 = 0$: